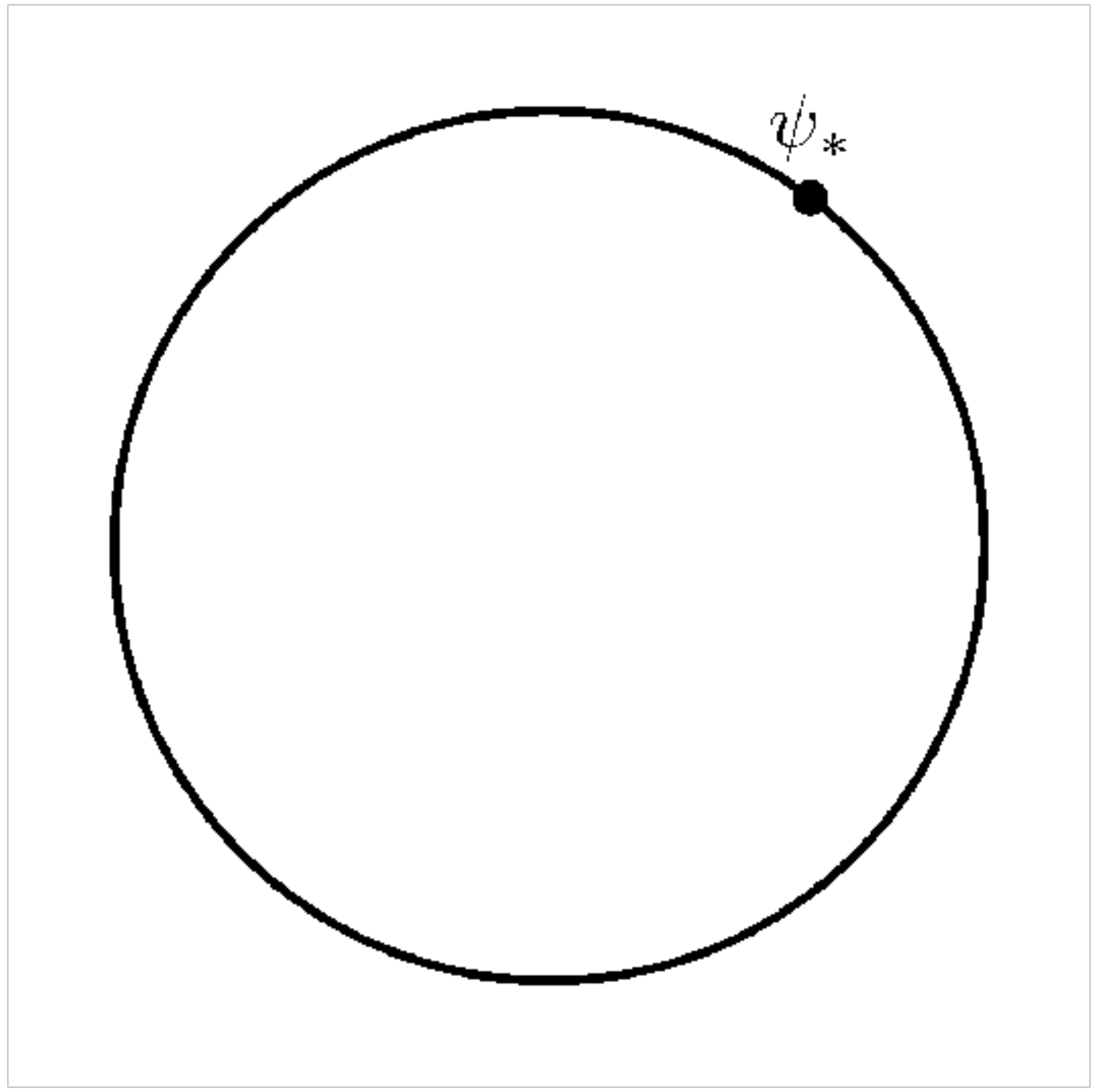


Probabilistic modeling with tensor networks

John Terilla

https://arxiv.org/abs/1902.06888

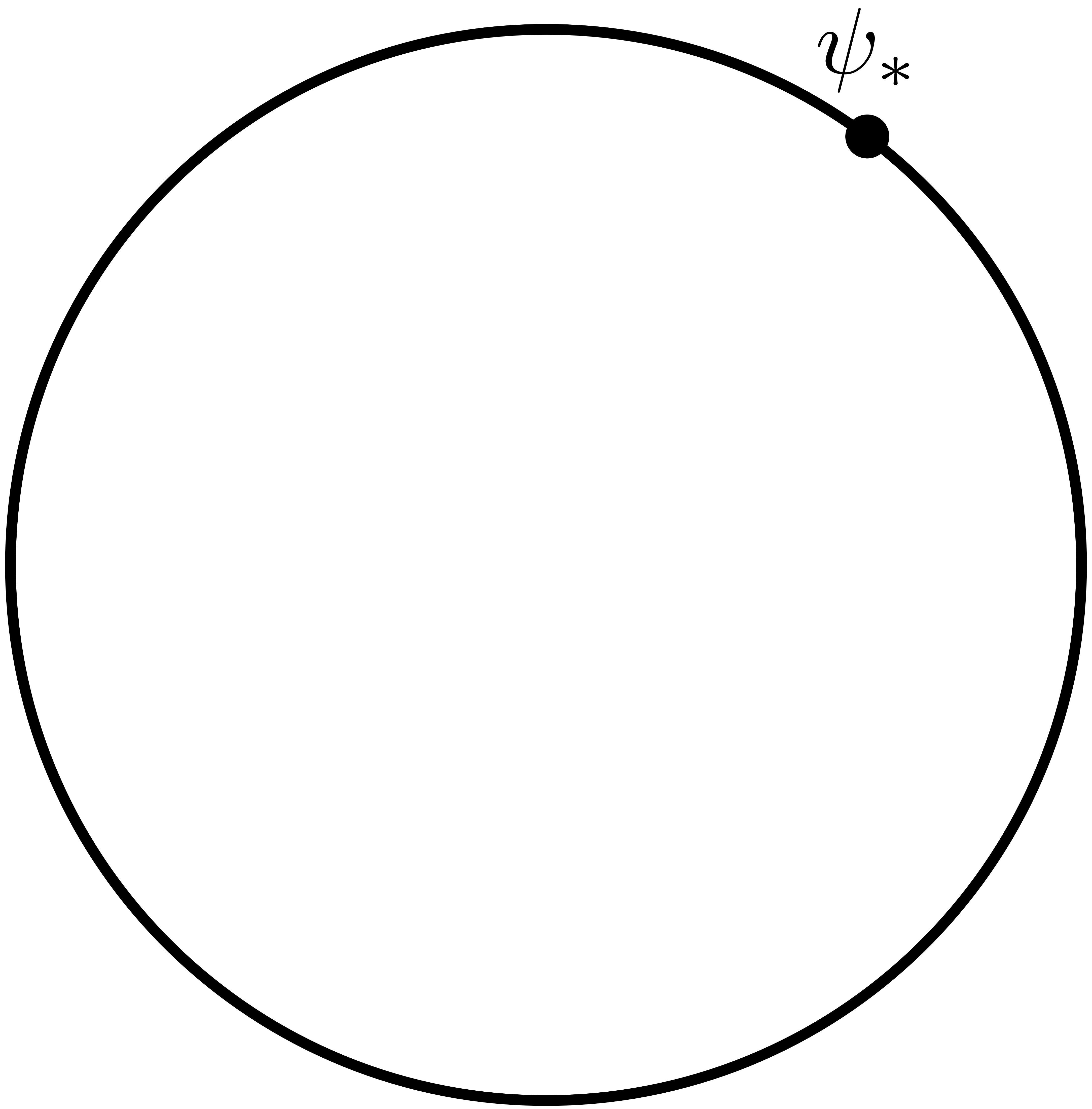
Joint with James Stokes at Tunnel in NY

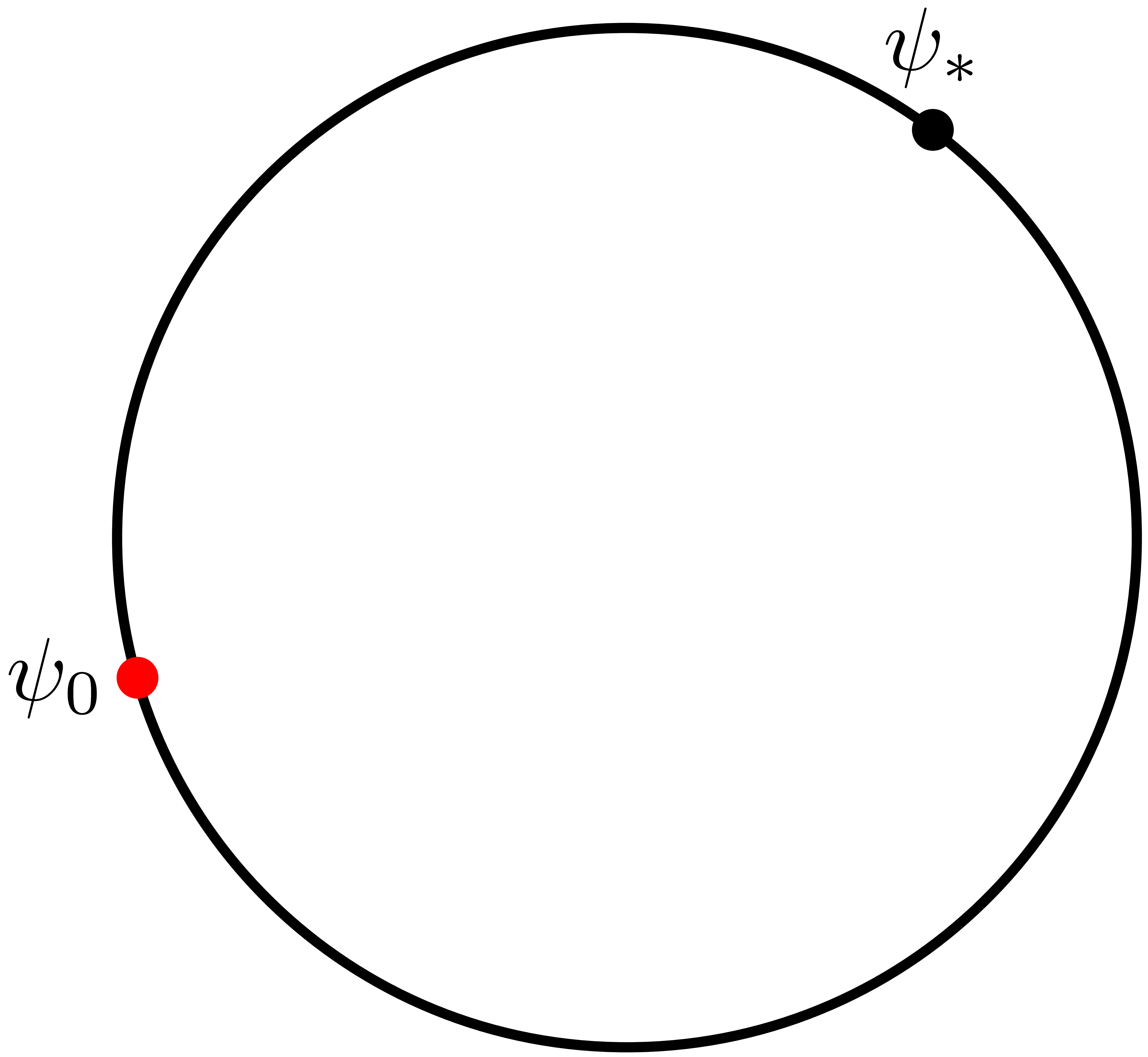
Quantum circuits are expressive.
Do they provide a useful inductive bias for machine learning?

We address this question in the context of unsupervised learning of probability distributions on sets of sequences.

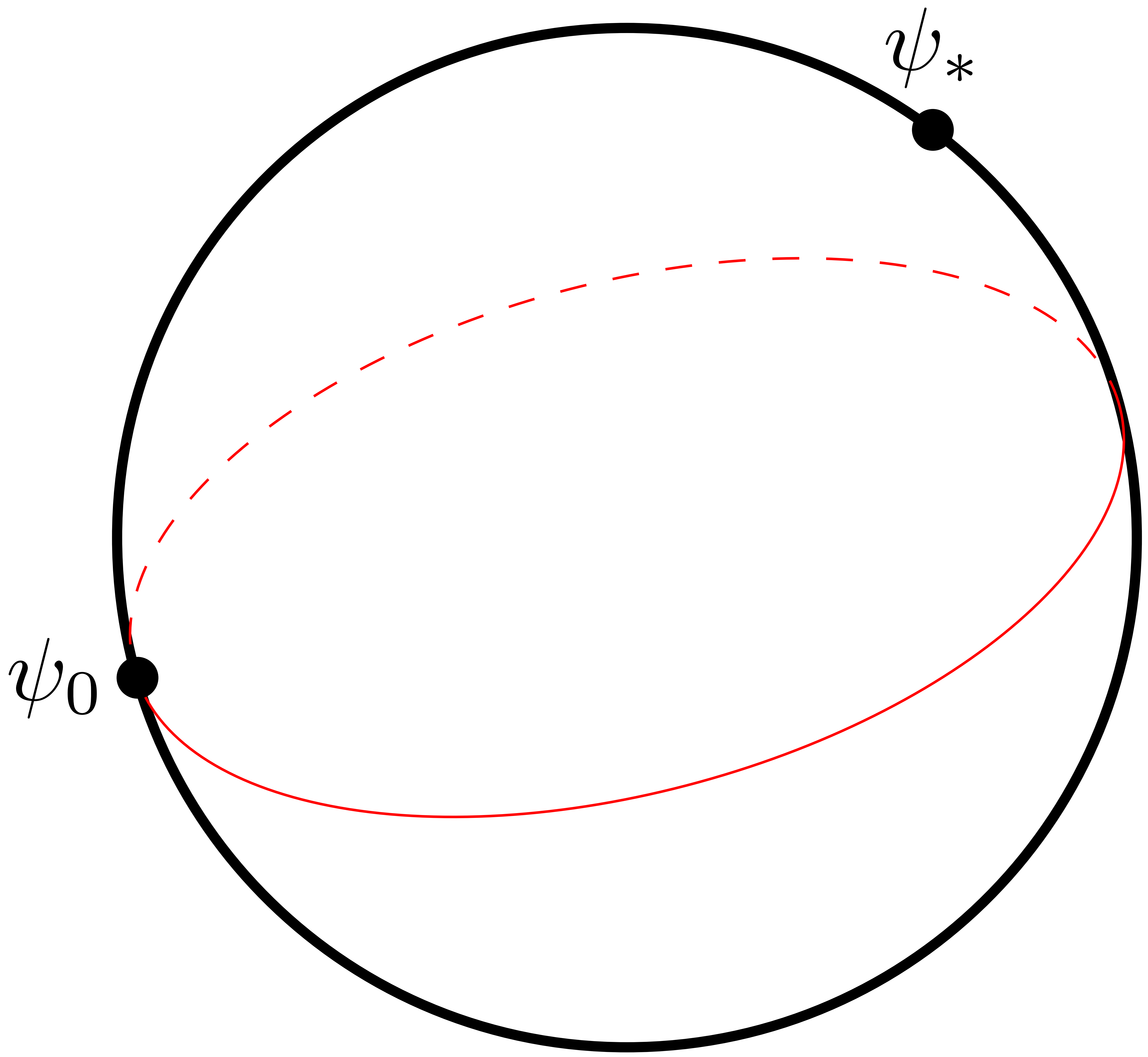
Idea: View sequences as observations of a one dimension system of interacting quantum particles. Then find the state of that system.

Use exact-DMRG to find the state

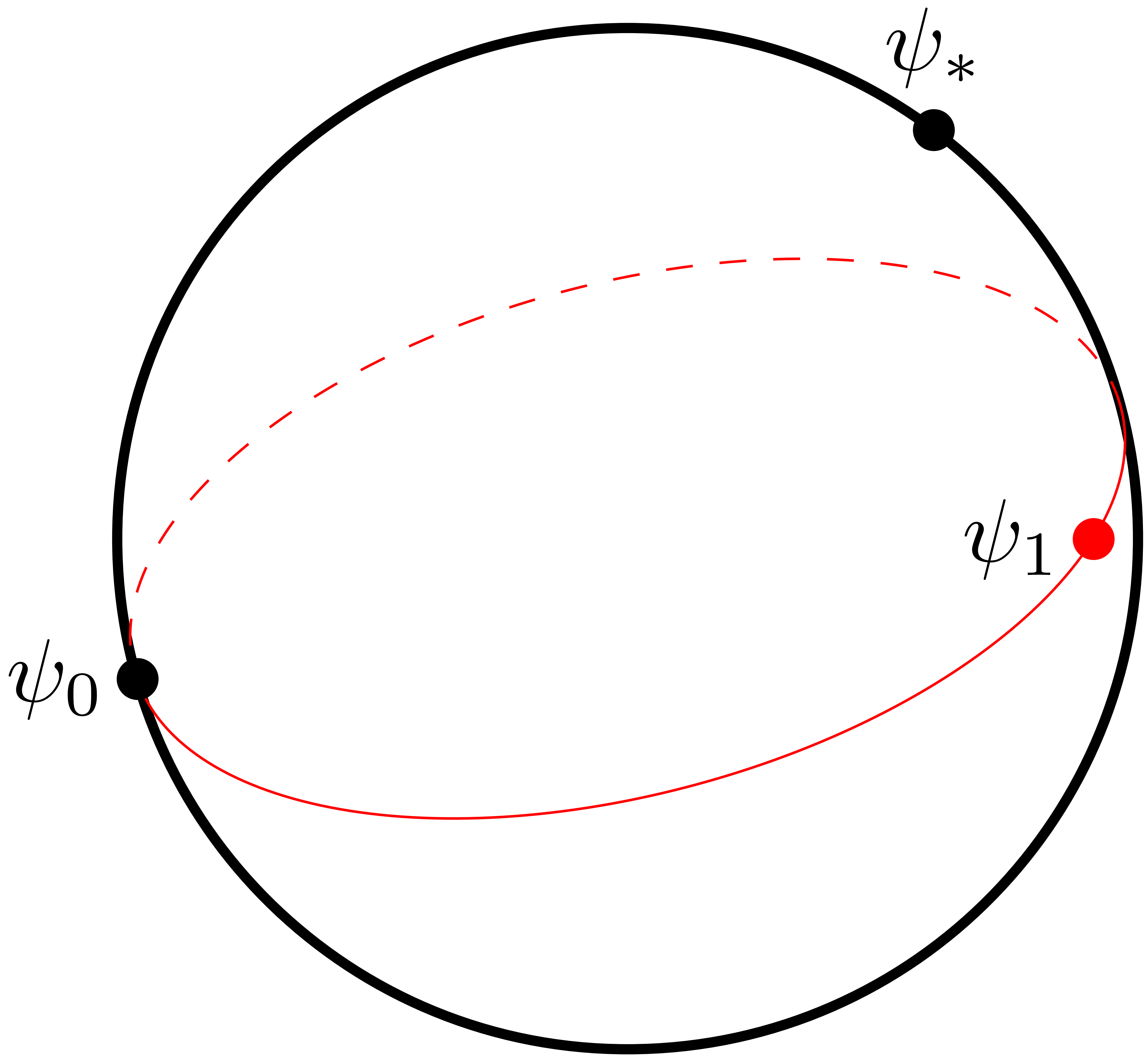

ψ_*

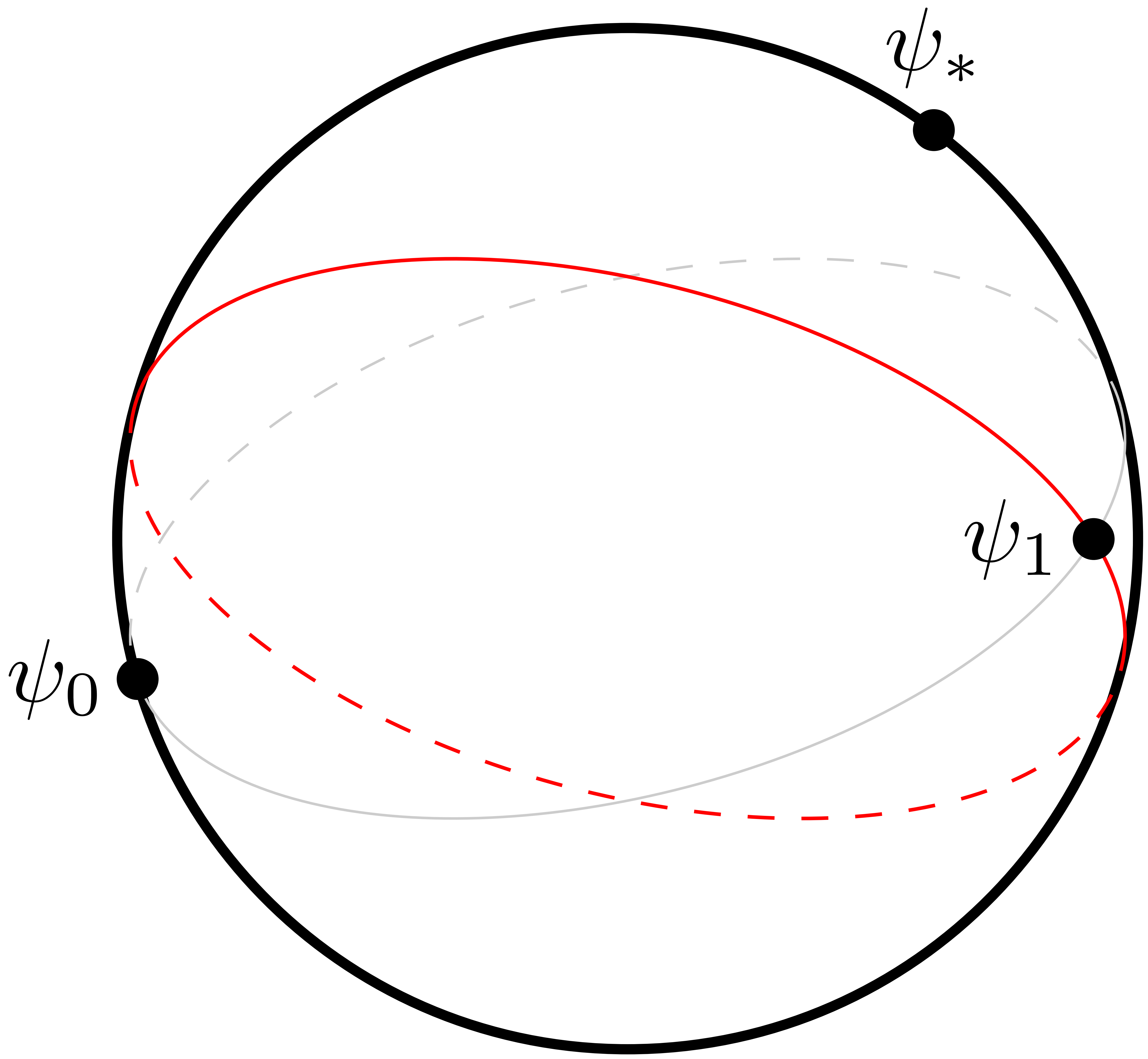


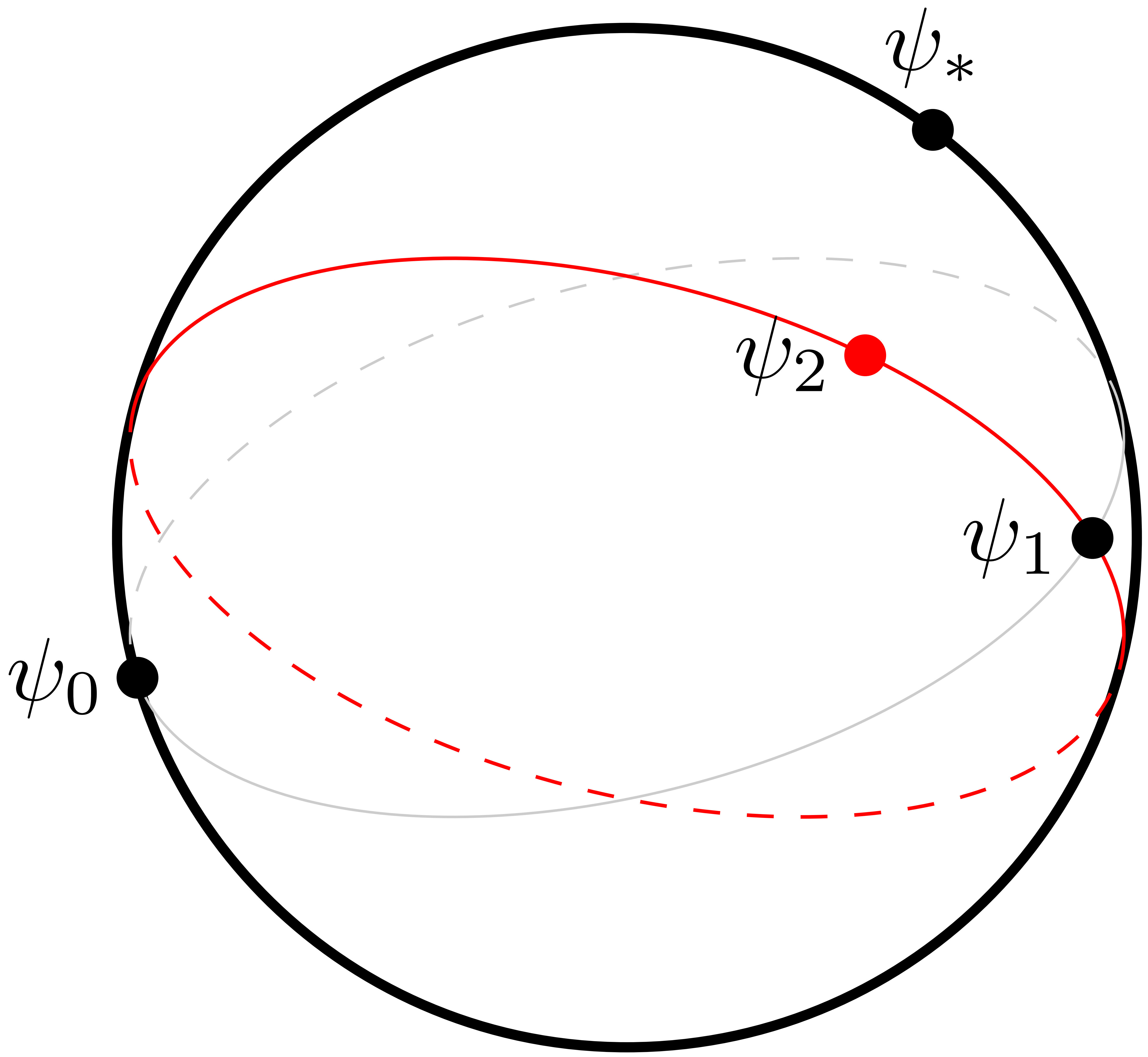
ψ_*
ψ_0

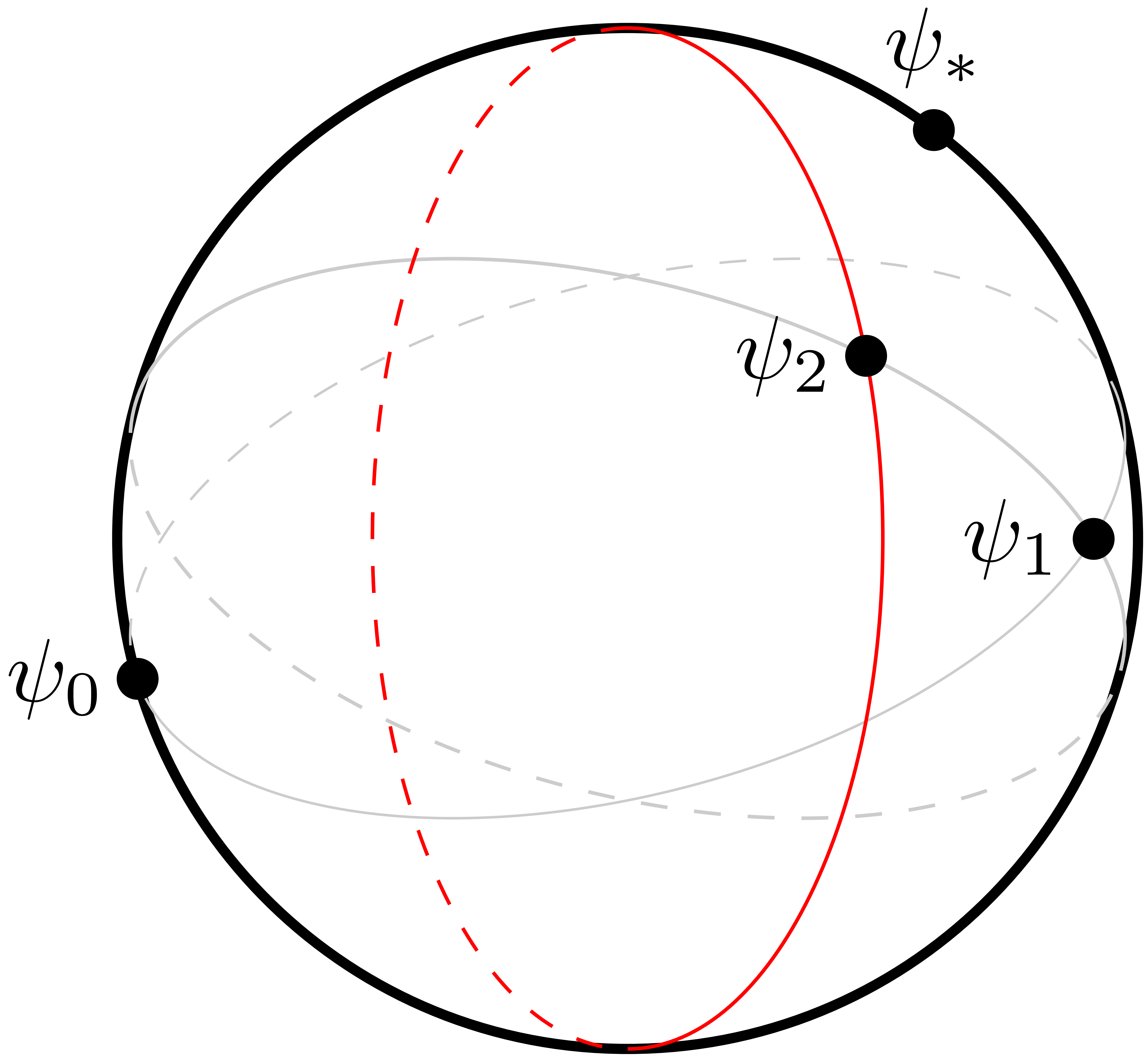


ψ_*
ψ_0

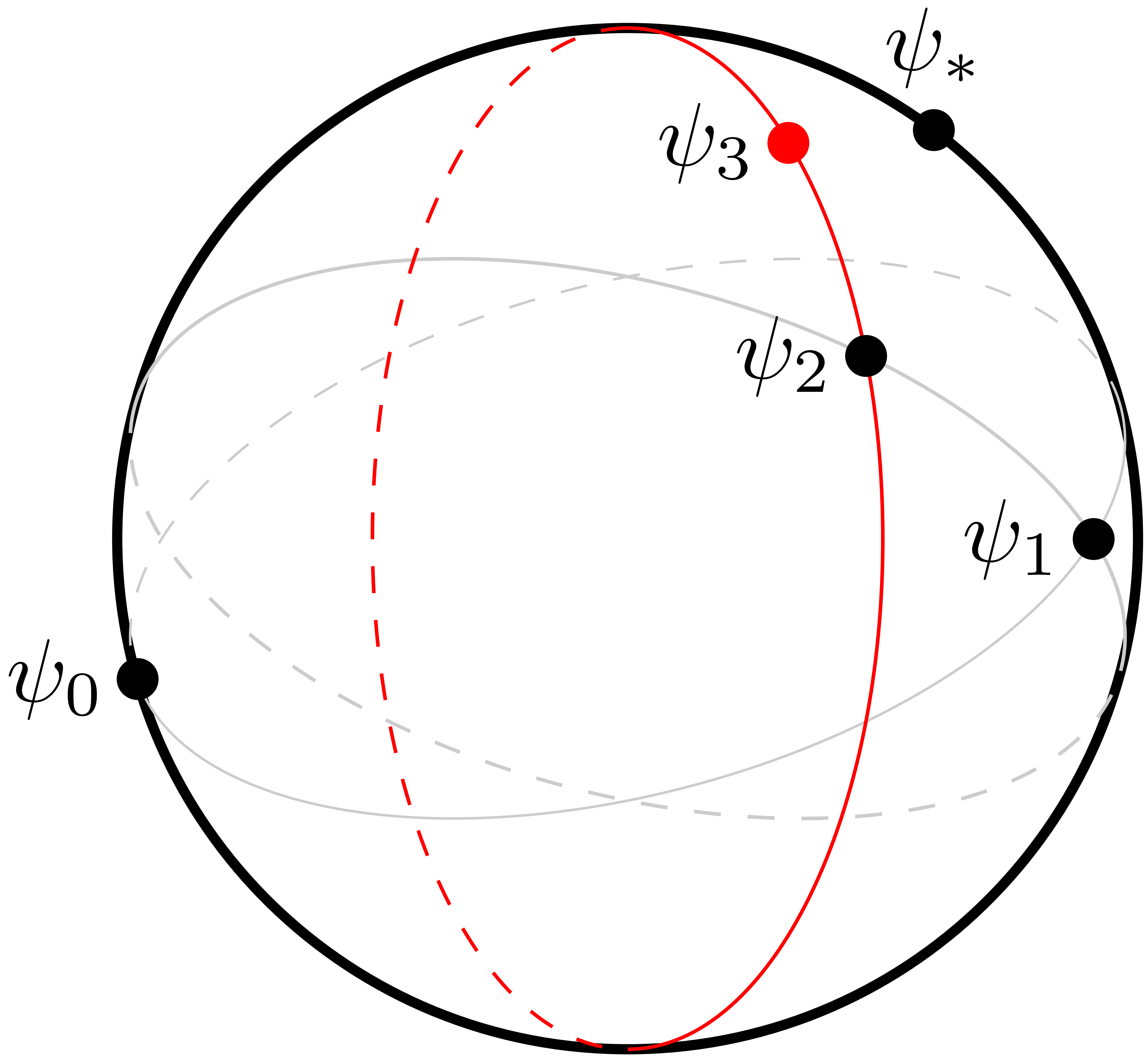

ψ_*
ψ_1
ψ_0


ψ_*
ψ_1
ψ_0

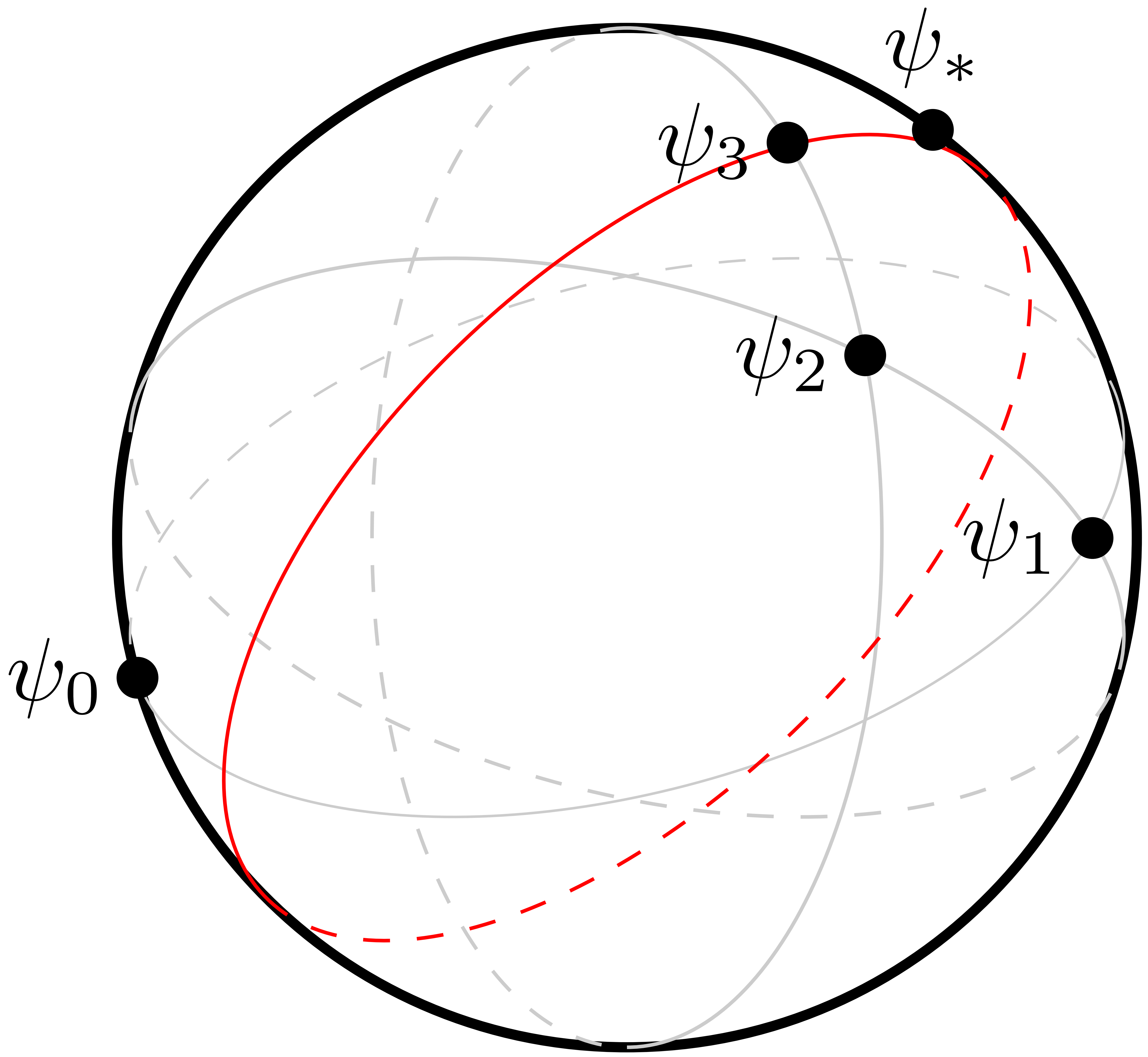

ψ_*
ψ_2
ψ_1
ψ_0



ψ_*
ψ_2
ψ_1
ψ_0



ψ_*
ψ_3
ψ_2
ψ_1
ψ_0


ψ_*
ψ_3
ψ_2
ψ_1
ψ_0

Dive into details

Let X be a finite set and consider $H = \mathbb{C}^X$, the free vector space on X.

Notation: for any $x \in X$, we have $|x\rangle \in H$

The space H has an inner product making it into a Hilbert space:

$$\langle x, x' \rangle = \delta_{x,x'}$$

Begin with a set X of sequences

$$X = A \times A \times \cdots \times A$$

and a probability distribution π on X.

The free Hilbert space $H = \mathbb{C}^X$ decomposes as a tensor product

$$H \simeq V \otimes V \otimes \cdots \otimes V$$

where $V = \mathbb{C}^A$ is the free Hilbert space on A.

Notation: $|a_1 a_2 \cdots a_N\rangle = |a_1\rangle \otimes |a_2\rangle \otimes \cdots \otimes |a_N\rangle$

The state $\psi_* = \sum_{x \in X} \sqrt{\pi(x)} |x\rangle$ encodes the probability distribution π via the Born rule $\pi(x) = \langle x | P_{\psi_*} | x \rangle$

Problem formulation

Given a set of samples drawn from π

$$
\begin{aligned}
x^1 &= a_1^1 a_2^1 \cdots a_N^1 \\
x^2 &= a_1^2 a_2^2 \cdots a_N^2 \\
&\vdots \\
x^n &= a_1^n a_2^n \cdots a_N^n
\end{aligned}
$$

and a model hypothesis class $\mathcal{M} \subset H$,

find the state $\psi \in \mathcal{M}$ closest to ψ_*.

Method of attack

We have the empirical distribution $\widehat{\pi}$ defined by the data $x^1, \ldots, x^n$ and the corresponding empirical state

$$\widehat{\psi} := \sum_{x \in X} \sqrt{\widehat{\pi}(x)} |x\rangle.$$

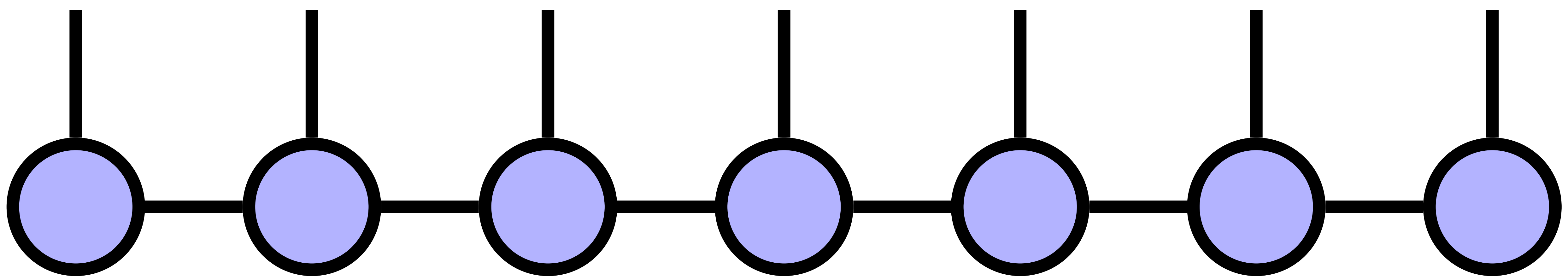
Use exact DMRG to define a sequence

$$\psi_0, \psi_1, \psi_2, \ldots$$

in $\mathcal{M}$ that get closer to $\widehat{\psi}$.

The model class $\mathcal{M}$

The model hypothesis class $\mathcal{M}$ consists of matrix product states (MPS) with a fixed bond space W.

$\psi =$ 

The model class $\mathcal{M}$

Tensor networks represent states that can be prepared by shallow quantum circuits.

Allow for efficient representations of vectors in very high dimensional spaces

Provide access to poly-logarithmic algorithms for certain kinds of linear algebra operations

Exact DMRG as a sequence of inductively defined effective problems

Base step: Choose an MPS state ψ_0

Inductive step: Given an MPS state ψ_t, define an isometric embedding of an "effective" Hilbert space

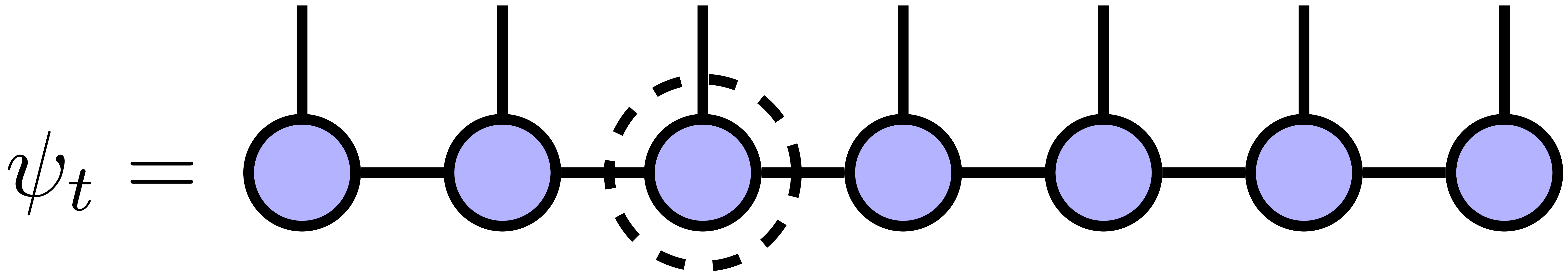
$$\alpha_t : H_{\mathrm{eff}} \to H.$$

Define ψ_{t+1} to be the state in $H_{t+1} := \alpha_t(H_{\mathrm{eff}})$ closest to $\widehat{\psi}$.

The state in $\alpha(H_{\text{eff}})$ that is closest to $\widehat{\psi}$ can be computed directly using orthogonal projection onto the subspace $\alpha(H_{\text{eff}})$

$$\begin{aligned}
\operatorname{Proj}(\widehat{\psi}) &= \alpha\alpha^* \left(\widehat{\psi}\right) \\
&= \alpha\alpha^* \left(\sum_{x\in X} \sqrt{\widehat{\pi}(x)}\, |x\rangle\right) \\
&= \alpha \left(\sum_{x\in X} \sqrt{\widehat{\pi}(x)}\, \alpha^* (|x\rangle)\right).
\end{aligned}$$

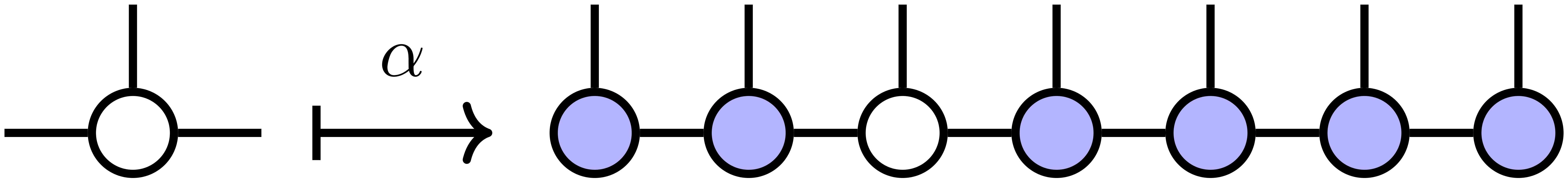
To define the effective problem, fix a site and put the MPS into mixed canonical gauge relative to that site.



The effective Hilbert space is $W \otimes V \otimes W$ and the isometric embedding

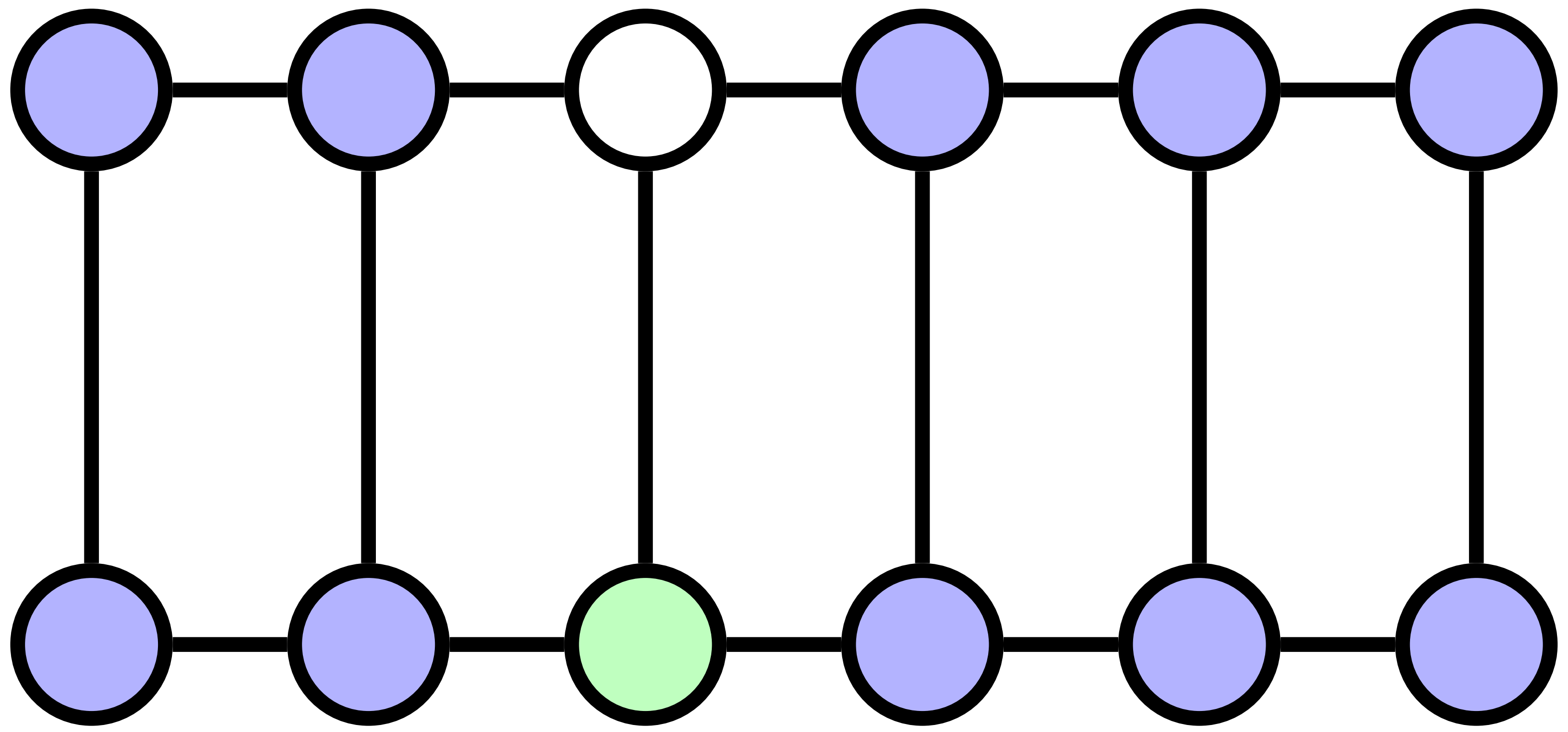
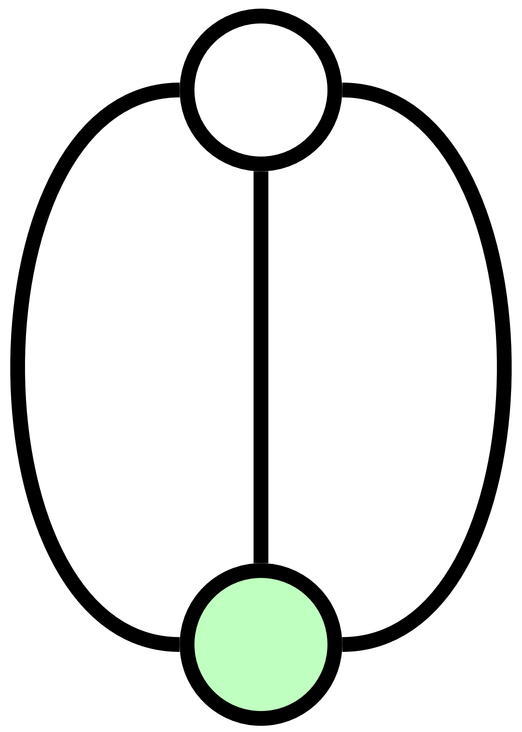
$$\alpha : W \otimes V \otimes W \to V^{\otimes N}$$

is defined by

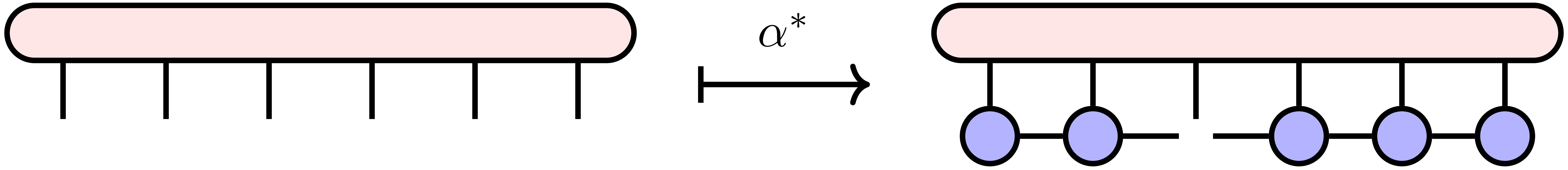


The map α is an isometry.

Proof:

$\langle \alpha(\phi), \alpha(\phi') \rangle =$  $=$  $= \langle \phi, \phi' \rangle.$

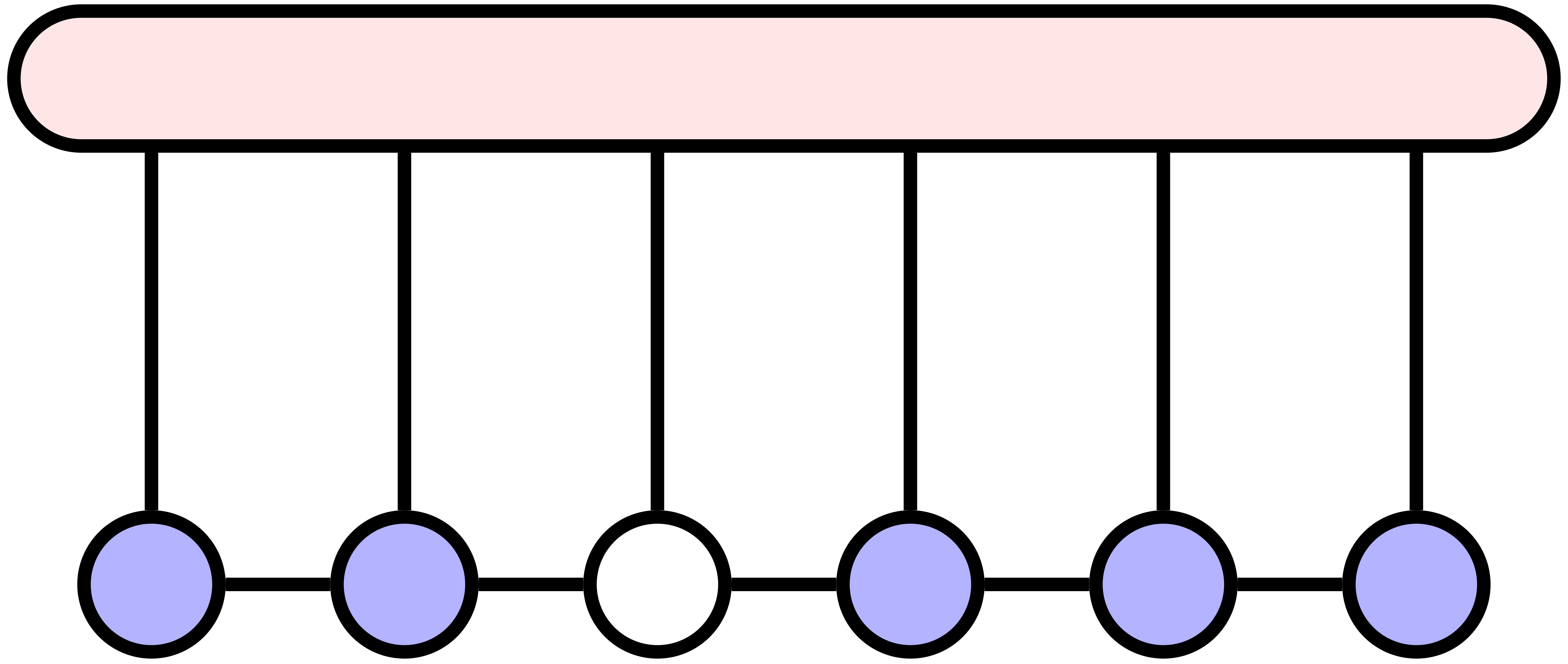
Define a map $\alpha^* : V^{\otimes N} \to W \otimes V \otimes W$ by



The map α^* is the adjoint of α.

Proof:

$$\langle \eta, \alpha(\phi) \rangle =$$

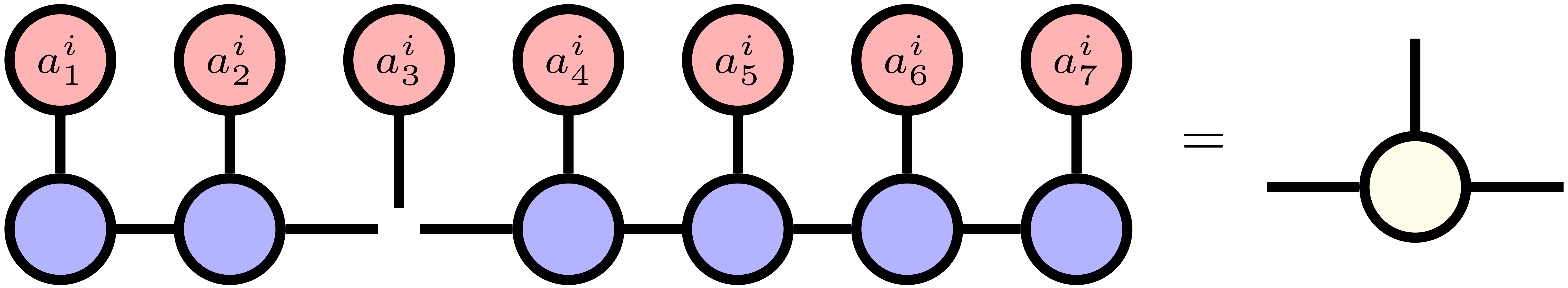


$$= \langle \alpha^*(\eta), \phi \rangle$$

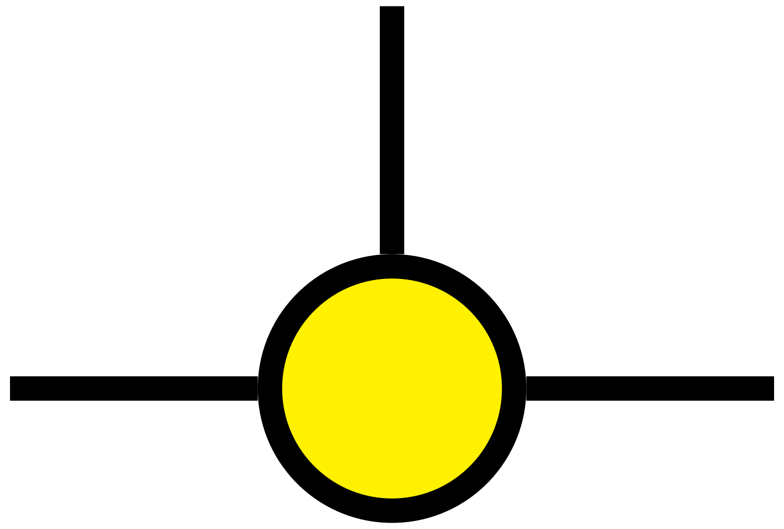
The state in $\alpha(H_{\text{eff}})$ that is closest to $\widehat{\psi}$ can be computed directly using orthogonal projection onto the subspace $\alpha(H_{\text{eff}})$

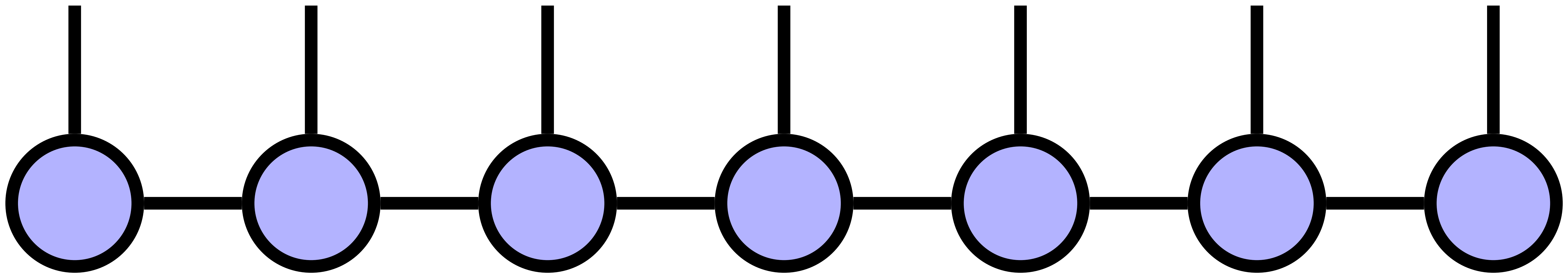
$$\begin{aligned}
\operatorname{Proj}(\widehat{\psi}) &= \alpha\alpha^* \left(\widehat{\psi}\right) \\
&= \alpha\alpha^* \left(\sum_{x \in X} \sqrt{\widehat{\pi}(x)}\, |x\rangle\right) \\
&= \alpha \left(\sum_{x \in X} \sqrt{\widehat{\pi}(x)}\, \alpha^*(|x\rangle)\right).
\end{aligned}$$

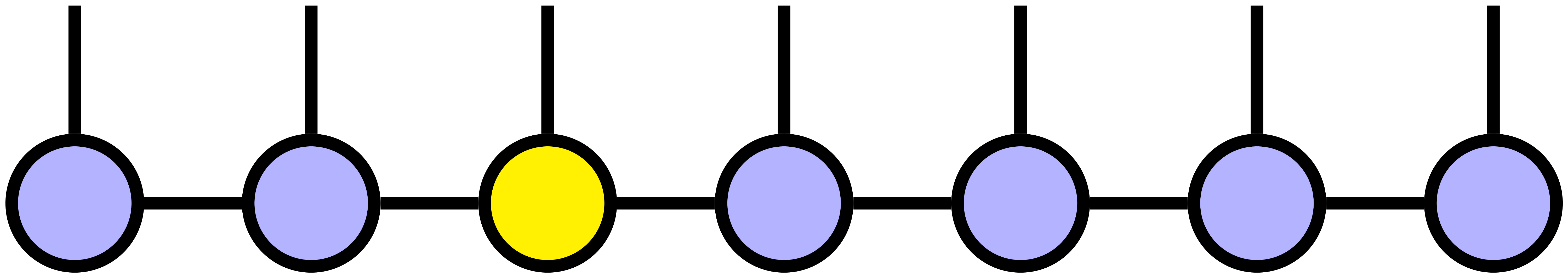
$$\alpha^* \left(|x^i\rangle \right) =$$

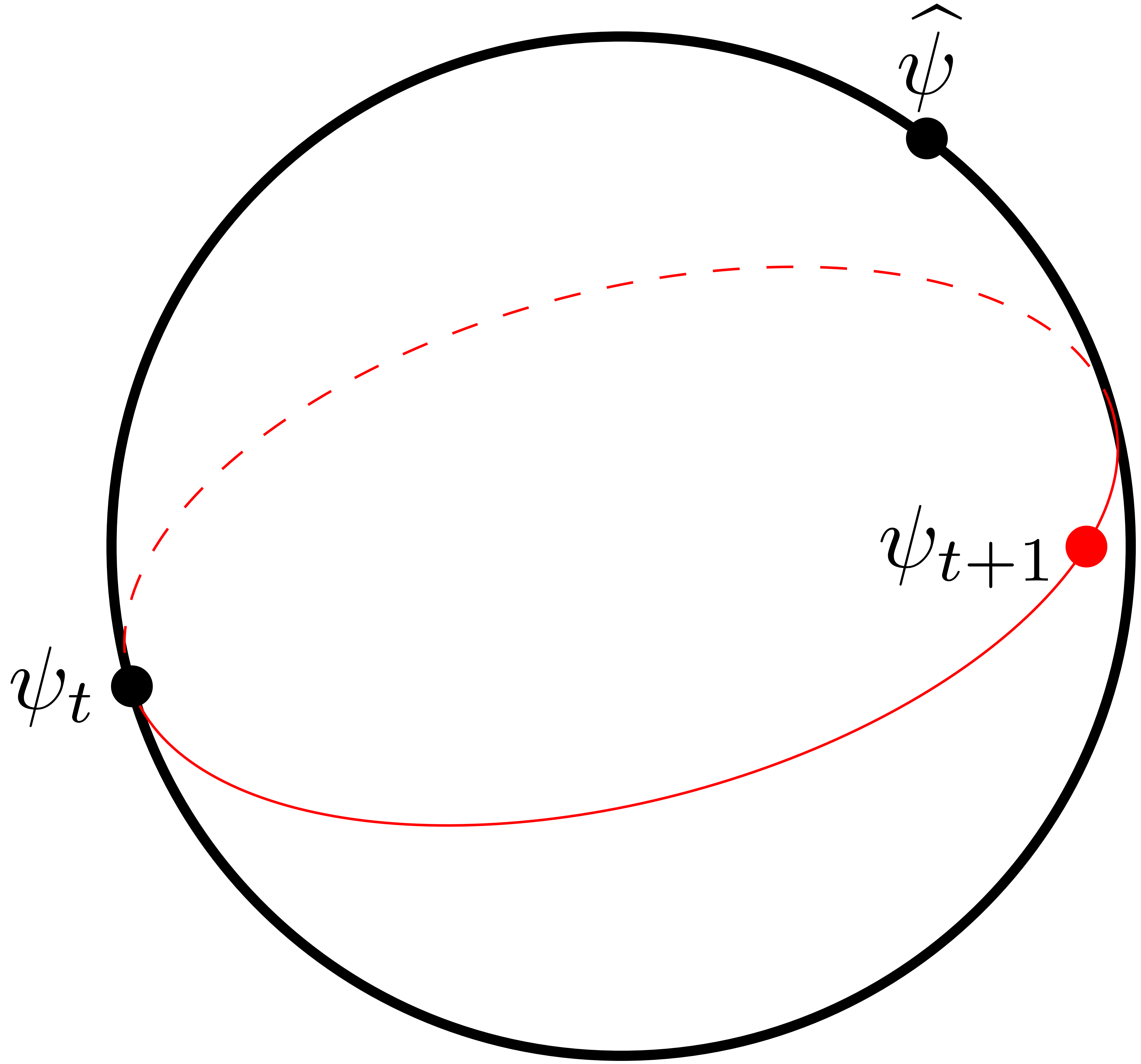


$$\sum_{i=1}^{n} c_i \alpha^* \left(|x^i\rangle \right) =$$



$\psi_t =$ 

$\psi_{t+1} =$ 



$\widehat{\psi}$
ψ_{t+1}
ψ_t

Experimental Results

Experimental Results

Parity Dataset P_N consists of bitstrings of length N with an even number of 1 bits

$$1100000101 \in P_{10}$$

$$1000000101 \notin P_{10}$$

Consider the probability distribution on bitstrings uniformly concentrated on P_N

$$\pi(1100000101) = \frac{1}{512}$$

$$\pi(1000000101) = 0$$

Experimental Results

Learns the uniform distribution on P_{20} with high accuracy

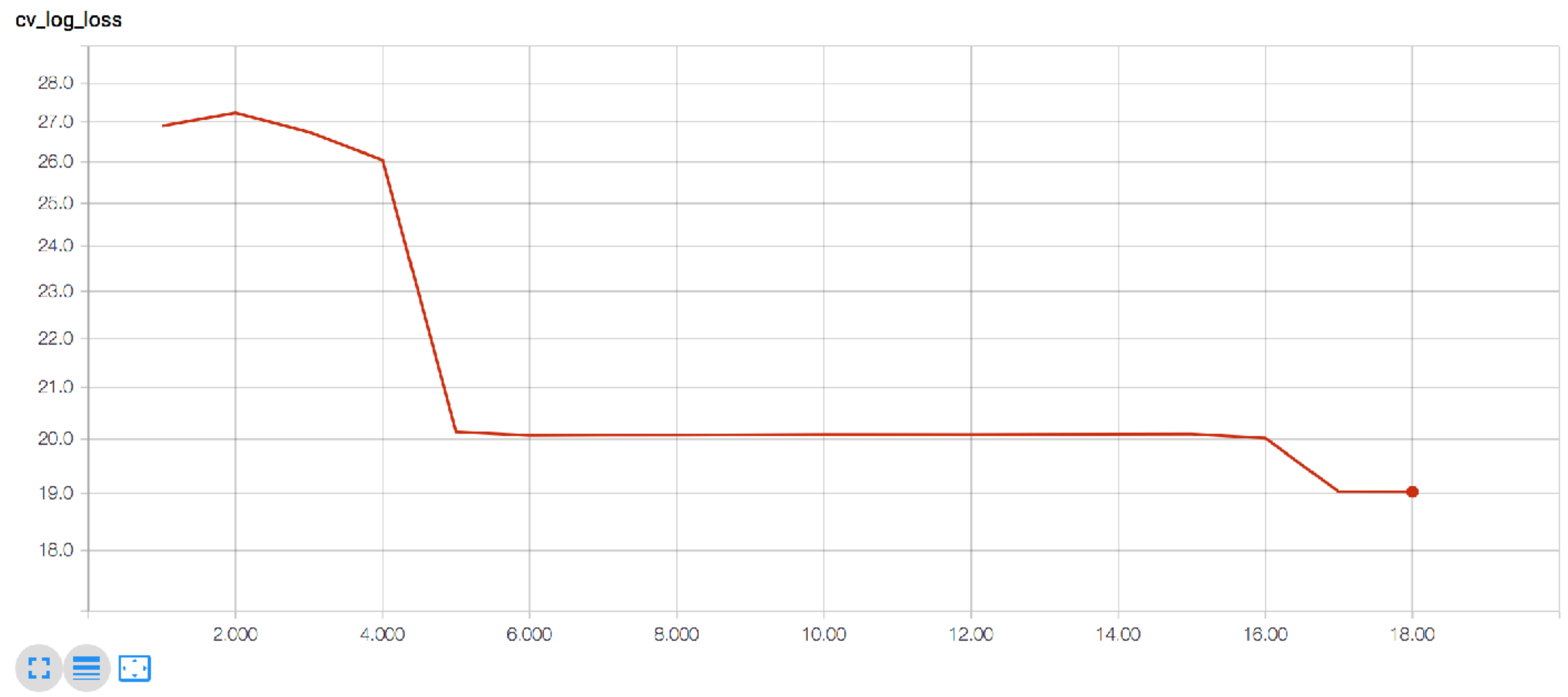
Used 2% of the data set to train

Trains quickly

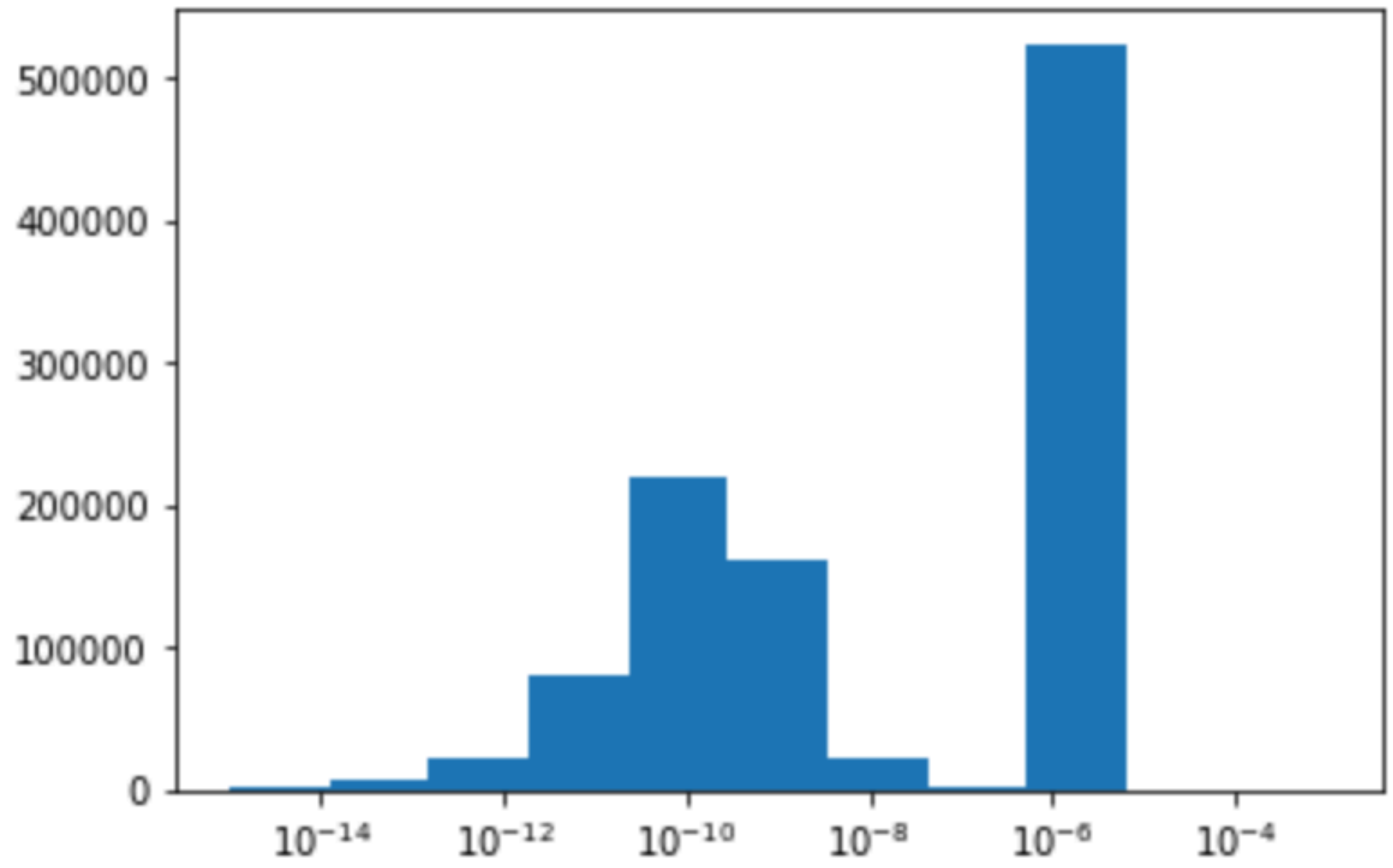
Resulting model is small (336 parameters)

Efficient, perfect sampling

Experimental Results



Experimental Results



Conclusions

The tensor network ansatz provides a useful inductive bias for unsupervised generative learning of datasets of interest

Other experimental results: DIV7

These methods could lead to interesting generative language models